Codage audio scalable basé sur le codeur MPEG-4 SSC

David VIRETTE[1], Jean-Bernard RAULT[2], Pierrick PHILIPPE[2]

France Telecom, Division R&D/TECH
prénom.nom@orange-ft.com
[1] : Laboratoire SSTP ; 2, Av. Pierre Marzin. 22307 Lannion Cedex France
[2] : Laboratoire IRIS ; 4, rue du Clos Courtel – BP 59 – 35512 Cesson Sévigné Cedex France

Résumé

Dans cet article, nous présentons un codeur audio scalable basé sur le codeur paramétrique MPEG-4 SSC (SinuSoidal Coder). Ce nouveau codeur combine deux stratégies de codage, la première étant le codage audio sinusoïdal (MPEG-4 SSC) et la deuxième étant le codage de type ACELP (Algebraic Code-Excited Linear Prediction), habituellement utilisé pour les signaux de parole. Nous montrons que cette approche permet d'une part, d'améliorer la qualité audio des codeurs paramétriques (sinusoïdaux) à bas-débits et d'autre part, d'offrir une flexibilité en terme de compromis qualité/débit comparé aux codeurs audio traditionnels.

Mots clefs

Codage audio paramétrique, codage sinusoïdal, MPEG4-SSC, scalabilité, ACELP.

1 Introduction

Depuis l'introduction du CD dans les années 80, et plus récemment avec l'explosion de l'Internet, les besoins en compression des signaux audio se sont rapidement développés. Les codeurs audio développés et standardisés par ISO/MPEG, comme le MP3, l'AAC ou l'HE-AAC présentés dans [1] et [2], sont largement utilisés de nos jours pour des applications de diffusion et de téléchargement de signaux audio. Ces algorithmes de codage audio, qui appartiennent à la famille des codeurs par transformée, exploitent les caractéristiques du système auditif humain, et notamment les effets de masquage fréquentiel, afin de réduire au maximum la distorsion perçue par l'auditeur sous contrainte de débit.

De nouvelles techniques de codage audio, généralement appelées codage audio paramétrique, ont été proposées plus récemment. Ces techniques s'appuient sur une décomposition du signal audio selon un modèle de codage simulant la façon dont le son est produit. Le signal audio est découpé en trames (quelques ms) pour être analysé relativement au modèle choisi. Les paramètres du modèle sont alors extraits, quantifiés et codés pour être transmis ou stockés. Au décodeur, le signal est re-synthétisé à l'aide des paramètres reçus. Citons en particulier le modèle sinusoïdal qui permet de modéliser les signaux audio à l'aide de simples oscillateurs, dont les paramètres (amplitudes, fréquences et phases) varient lentement dans le temps. Ces modèles ont été initialement développés dans les années 80 pour coder la parole en bande téléphonique [3].

Ces schémas d'analyse/synthèse ont ensuite été généralisés à tout type de signaux audio notamment avec le modèle Sinusoïdes + Bruit [4]. Dans ce modèle, le signal résiduel, obtenu une fois les composantes sinusoïdales retirées, est modélisé par un processus stochastique (bruit blanc) mis en forme temporellement et fréquentiellement.

Plus récemment, des modèles Sinusoïdes + Transitoires + Bruit ont été proposés afin d'améliorer la représentation des signaux percussifs [5].

Des algorithmes de codage audio ont été développés sur chacun de ces modèles. Nous pouvons citer par exemple le codeur MPEG-4 HILN (Harmonic and Individual Lines and Noise) [6] ou encore le codeur MPEG-4 SSC (SinuSoidal Coder) [7]. La qualité de ces codeurs paramétriques souffre d'un manque de naturel de par la limitation du nombre de composantes sinusoïdales sélectionnées et surtout par l'utilisation d'un simple modèle stochastique du résiduel.

Dans cet article, nous commencerons par présenter brièvement le codeur audio paramétrique MPEG-4 SSC en nous intéressant plus particulièrement à la modélisation du signal résiduel. Ensuite, nous proposerons un nouveau codeur audio basé sur l'association du codeur SSC et du codage ACELP. Nous verrons comment cette nouvelle structure de codage offre une plus grande flexibilité en termes de débit. Finalement, nous comparerons les performances de la solution proposée en comparaison avec le codeur MPEG-4 SSC. Cette comparaison sera effectuée à l'aide d'une mesure objective de qualité et par la réalisation d'un test subjectif formel.

2 Le codage audio paramétrique

Cette section présentera le standard MPEG-4 SSC qui est l'état de l'art en matière de codage audio paramétrique,

puis nous donnerons les différents points faibles du modèle utilisé.

2.1 Le codeur MPEG-4 SSC (SinuSoidal Coder)

Le codeur MPEG-4 SSC s'appuie sur un modèle Sinusoïdes + Transitoires + Bruit. Ce codeur fonctionne en bande HiFi à 44.1 kHz de fréquence d'échantillonnage. Les différentes composantes sonores de ce modèle sont représentées de la façon suivante :

- transitoires : sinusoïdes contraintes par une enveloppe temporelle;
- sinusoïdes : sinusoïdes contrôlées en amplitude, phase et fréquence;
- bruit : bruit aléatoire large bande mis en forme temporellement et spectralement.

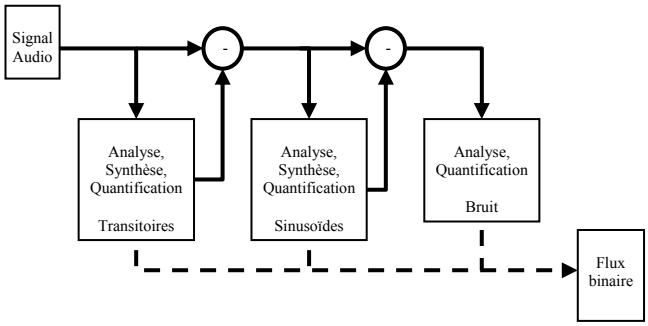

Figure 1 – *Codeur MPEG-4 SSC*

La figure 1 donne le schéma fonctionnel du codeur SSC. Les paramètres du modèle sont extraits en trois étapes successives. Tout d'abord, les transitoires sont détectées en mesurant les variations rapides et importantes de l'énergie du signal, puis modélisées et soustraites du signal original. Ensuite sur le signal restant, les composantes tonales qui sont perceptivement les plus importantes sont détectées et modélisées par des sinusoïdes, puis soustraites. Enfin, le signal déduit des deux étapes précédentes est considéré comme une composante de bruit. Il est modélisé par son enveloppe temporelle et fréquentielle. Les paramètres issus de ces trois étapes de codage sont ensuite quantifiés et multiplexés dans un flux binaire pour la transmission.

Le décodeur réalise les opérations de décodage et de synthèse des trois composantes du modèle afin de générer un signal perceptivement proche du signal original.

Nous allons nous intéresser plus particulièrement au module de synthèse de bruit décrit à la Figure 2. Comme le montre cette figure, la composante « Bruit » est synthétisée par un bruit large bande. Ce bruit est tout d'abord mis en forme temporellement à partir d'une enveloppe temporelle transmise sous forme de LSFs (Line Spectral Frequencies) [8] et convertie dans le domaine temporel. Le bruit est ensuite ajusté en énergie par des gains, également transmis. Les paramètres étant transmis trame par trame, un module de fenêtrage et d'Overlap-Add est ensuite utilisé pour la reconstruction du signal. Enfin, ce bruit est mis en forme spectralement par un filtre de Laguerre, qui offre une bonne résolution fréquentielle dans les basses fréquences [9].

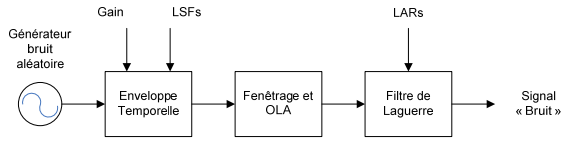

Figure 2 – *Synthétiseur du « Bruit » du décodeur MPEG-4 SSC*

Lors des tests de vérification réalisés par MPEG, le codeur MPEG-4 SSC a obtenu des notes MUSHRA entre *Convenable* et *Bonne* pour un débit de 24 kbps en stéréo [10].

2.2 Limites du codage audio paramétrique

Un problème bien connu concernant les modèles Transitoires + Sinusoïdes + Bruit est que, en général, le « bruit » résiduel n'est pas réellement un bruit et ceci pour les raisons suivantes :

- Le nombre limité de sinusoïdes transmises implique que le signal résiduel peut encore contenir des composantes tonales;
- Les paramètres des sinusoïdes (amplitude, phase et fréquence) peuvent avoir été mal estimés. La soustraction de ces composantes, quantifiées, peut entrainer la présence de caractéristiques tonales dans le résiduel;

De plus certains signaux audio ne sont pas adaptés au modèle (nombre fini de sinusoïdes), ce qui implique que le résiduel est fortement « coloré » et donc mal modélisé par un processus stochastique. La qualité des codeurs paramétriques souffre en général d'un manque de réalisme. Des informations de localisation ou d'ambiances sont souvent éliminées, ce qui entraine, en général, un manque de « présence » et de « naturel ».

Une conséquence importante de ces deux dernières limitations est que, même en augmentant le débit associé à ce codeur, la transparence ne peut être atteinte. En se basant sur ces limites du codage audio paramétrique, nous proposons donc une nouvelle architecture afin d'en améliorer la qualité.

3 SSC-ACELP scalable

3.1 Codeur

Ayant présenté le codeur SSC et les limitations associées au modèle Sinusoïdes + Transitoires + Bruit, nous allons considérer une nouvelle architecture de codage associant le codeur SSC avec un codage ACELP en sous-bande comme le montre la Figure 3. Dans cette nouvelle architecture de codage, le codeur SSC, tel que décrit dans la section précédente, est utilisé comme codeur principal. Ensuite, un codage du résiduel SSC est réalisé en sous-bandes, suivant ainsi les principes psychoacoustiques basé sur la sensibilité de l'oreille humaine en fréquence. Cette découpe en sous-bandes permet de mieux répartir le débit des sous-codeurs sur chaque sous-bande. Il est ainsi possible d'associer un débit plus élevé aux sous-bandes basses, qui seront confiées à un codage ACELP. Pour les hautes fréquences, le module de synthèse de bruit sera souvent suffisant pour assurer un codage de bonne qualité. On le voit donc, le débit peut être consacré à représenter efficacement les premières bandes qui sont les plus significatives perceptivement.

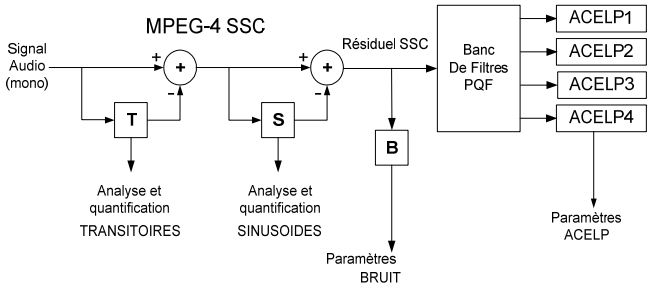

Figure 3 – *Codeur MPEG-4 SSC + ACELP*

La séparation en quatre sous-bandes est réalisée par un banc de filtre de type PQF (Polyphase Quadrature Filter) utilisé dans le processus de contrôle de gain du MPEG-2 AAC [11]. Les coefficients du banc de filtres d'analyse sont donnés par la formule suivante :

$$h_i(n) = \frac{1}{4}\cos\left(\frac{(2i+1)(2n+5)\pi}{16}\right)Q(n)$$

$$0 \le n \le 95, 0 \le i \le 3$$

Avec $Q(n) = Q(95-n), 48 \le n \le 95$

Q(n) représente le filtre à réponse impulsionnelle finie (FIR) prototype passe-bas de longueur 96.

Le schéma de codage décrit à la Figure 3 permet de définir directement un format de flux binaire scalable associant des couches additionnelles de codage ACELP (bande 1 à 4) au cœur SSC.

Les modules de codage ACELP ont été adaptés à partir du codeur AMR-WB (Adaptive Multi-Rate – WideBand) normalisé au 3GPP comme codeur conversationnel en bande élargie [12]. La trame de l'AMR-WB est composée de 4 sous-trames de 64 échantillons. Pour chaque sous-trame, un filtre de prédiction linéaire, une excitation adaptative (pitch et gain) et une excitation algébrique (impulsions et gain) sont sélectionnées pour modéliser le signal. L'AMR-WB possède plusieurs débits de fonctionnement définis principalement par les tailles des dictionnaires algébriques utilisés. Ces dictionnaires sont imbriqués de par leur construction. Le débit d'un mode de l'AMR-WB est donc défini par le nombre d'impulsions +/-1 sélectionnées pour construire l'excitation algébrique.
Dans le codeur SSC-ACELP, les modules ACELP travaillent sur des trames composées de 6 sous-trames de 64 échantillons. Les modifications apportées à l'AMR-WB portent principalement sur la résolution de la recherche du pitch pour l'excitation adaptative (pitch entier uniquement) et sur la quantification des gains (2 gains quantifiés en absolu et 4 en relatif). Les différents débits du codeur SSC-ACELP sont définis par les dictionnaires algébriques sélectionnés dans chaque sous-bande.

3.2 Décodeur

Le décodeur associé est décrit à la Figure 4. Nous pouvons noter que dans un premier temps un décodage conforme au MPEG-4 SSC est réalisé. Le signal alors généré offre la qualité du codeur paramétrique. Ensuite, en fonction des couches ACELP reçues, les différentes sous-bandes préalablement décodées sont remplacées.

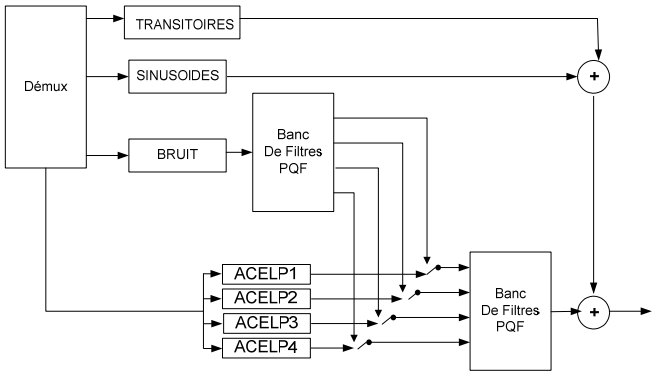

Figure 4 – *Décodeur scalable MPEG-4 SSC + ACELP*

La Figure 5 présente le train binaire associé au codeur SSC-ACELP. Dans le cas le plus simple, il comporte 5 couches permettant d'améliorer la qualité en remplaçant la synthèse de bruit du SSC dans une sous-bande par le décodage ACELP associé. Toutefois, cette structure de train binaire peut être enrichie de couches supplémentaires de raffinement des excitations algébriques des modules

ACELP. Ainsi, les différentes sous-bandes seront encodées par un ACELP multi-étage. Dans ce cas particulier, pour réduire la complexité, des méthodes de transcodage ACELP peuvent être exploitées à l'encodage lors de la recherche des codes algébriques [13]. On pourra par exemple effectuer la recherche dans le dictionnaire ACELP le plus riche afin de favoriser la qualité du débit le plus élevé, puis « dégrader » le code algébrique choisi en supprimant certaines impulsions afin qu'il reste compatible avec les débits plus faibles.

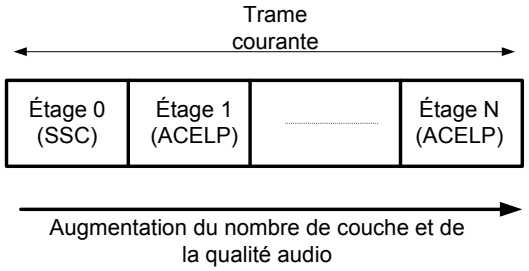

Figure 5 – *Flux binaire MPEG-4 SSC + ACELP pour la scalabilité*

Dans cette section, nous avons présenté la nouvelle architecture de codage proposée. En associant le codage audio paramétrique (SSC) au débit nominal de 24 kbps avec des modules de codage ACELP multi-étage, il est possible d'obtenir une granularité d'environ 4 kbps par couche entre 24 kbps et 128 kbps pour offrir une amélioration continue de la qualité perçue.

4 Performances

4.1 Test subjectif

Dans cette section, nous allons présenter les résultats d'un test subjectif qui a été réalisé dans le but d'évaluer la qualité audio de l'architecture de codage SSC-ACELP. Ce test visait à montrer que l'association du MPEG-4 SSC avec une seule sous-bande de codage ACELP améliore la qualité audio. Dans ce mode, le débit associé à la partie Sinusoïdes + Transitoires + Bruit est d'environ 18 kbps, alors que le débit associé à la première sous-bande ACELP est d'environ 6 kbps. Le module de codage ACELP de la première sous-bande de fréquence utilisé dans ce cas est le dictionnaire algébrique de plus faible débit (2 impulsions sur les 64 positions). Ce mode de codage est donc comparé au SSC à un débit de 24 kbps.
Le test d'écoute a été réalisé en suivant la méthodologie de test CMOS avec l'échelle de notation définie dans la Recommandation ITU-R BS.562-3. Selon cette méthodologie de test, pour chaque signal audio, l'ordre d'écoute est Ref/A/B, avec Ref correspondant au signal de référence (signal original dans notre cas), A et B sont les signaux à évaluer présentés dans un ordre aléatoire non connu du sujet (« test en aveugle »). Dans notre cas, A et B représentaient soit le signal audio encodé avec le MPEG-4 SSC, soit avec le SSC-ACELP, tous deux à un débit de 24 kbps. Les signaux audio de test étaient composés des 12 signaux critiques habituellement utilisés par MPEG pour l'évaluation des codecs audio. Cette liste est donnée à la Figure 6.

Item	Description
es01	vocal (Suzanne Vega)
es02	German speech
es03	English speech
si01	Harpsichord
si02	Castanets
si03	Pitch pipe
sm01	Bagpipes
sm02	Glockenspiel
sm03	Plucked strings
sc01	Trumpet solo and orchestra
sc02	Orchestral piece
sc03	Contemporary pop music

Figure 6 – *Liste des signaux de test*

Huit sujets ont participé à ce test. La Figure 7 montre les résultats de ce test pour chaque signal et en moyenne sur l'ensemble des 12 signaux. Cette figure montre le score moyen pour chaque signal, ainsi que l'intervalle de confiance à 95%. Il apparaît que le codage SSC-ACELP améliore la qualité sur les échantillons de parole (es01, es02 et es03) de manière significative. Par contre, sur les échantillons de musique, il n'y a pas de différence significative à débit équivalent. Ces résultats peuvent s'expliquer par le fait que la parole encodée par un codeur sinusoïdal manque de naturel. L'utilisation d'un schéma de codage ACELP permet donc d'améliorer la qualité sur ces signaux critiques.

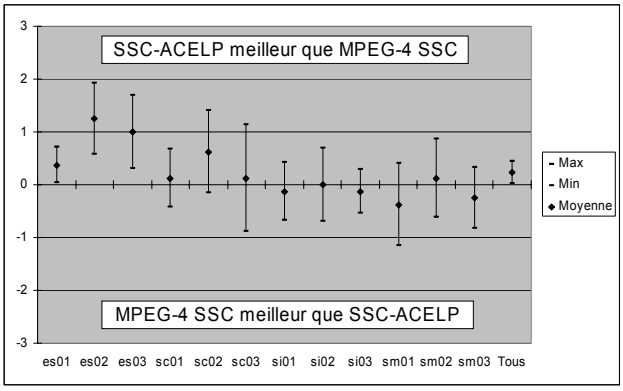

Figure 7 – *Résultats du test subjectif*

4.2 Mesure objective

Nous avons également utilisé l'outil PEAQ (Perceptual Evaluation of Audio Quality) pour mesurer les performances de la structure de codage proposée. Cet outil développé par l'ITU-R dans la recommandation BS-1387 [14] permet de fournir une note appelée ODG (Objective Difference Grade), représentative de la qualité audio du signal testé. Le résultat de l'ODG donne une note comprise entre 0 et -4, où 0 correspond à une dégradation imperceptible et -4 à une dégradation très gênante. La note est négative car le signal testé est considéré comme moins bon que le signal de référence.
Les notes ODG des échantillons encodés avec le codeur SSC-ACELP sont meilleures en moyenne que la référence MPEG. Nous pouvons aussi noter que pour la majorité des échantillons, le SSC-ACELP est meilleur que le codeur MPEG4-SSC. La figure 8 montre les résultats détaillés pour les deux codeurs.

Item	SSC-ACELP	MPEG4-SSC
es01	-2.359	-3.491
es02	-2.637	-3.418
es03	-2.636	-3.514
si01	-2.344	-2.827
si02	-3.609	-3.818
si03	-1.080	-1.960
sm01	-1.819	-2.362
sm02	-3.098	-2.444
sm03	-1.904	-3.362
sc01	-3.297	-3.261
sc02	-2.848	-3.549
sc03	-1.557	-3.388
Moyenne	-2.432	-3.116

Figure 8 – *Résultats ODG*

5 Conclusion

Dans cet article, nous avons introduit le codeur audio scalable MPEG-4 SSC-ACELP. Nous avons présenté l'intérêt de combiner le codage audio paramétrique avec des modules de codage ACELP. Cette nouvelle architecture de codage permet de mieux représenter le signal résiduel et offre une grande flexibilité (scalabilité) en termes de débit. Des tests subjectifs à 24 kbps ont montré que ce nouveau codeur permet d'offrir une meilleure qualité audio qu'un codeur audio paramétrique « état de l'art ». De nouvelles évaluations seront menées dans le but de caractériser les performances du codeur scalable à différents points de fonctionnement. Il sera ainsi intéressant de confirmer de manière formelle l'amélioration continue de la qualité constatée de manière informelle.

Références

[1] Karlheinz Brandenburg. "MP3 and AAC Explained", Présenté à la 17ème Conférence International AES, Florence, Italie, Septembre 1999.

[2] Martin Wolters, Kristofer Kjörling, Daniel Homm, Heiko Purnhagen, "A closer look into MPEG-4 High Efficiency AAC", 115ème Convention AES, New York, USA, Octobre 2003.

[3] R.J. McAulay et T.F. Quatieri, "Speech analysis & synthesis based on a sinusoidal representation", IEEE Trans. on ASSP, Vol. 34, No. 4, Août 1986.

[4] B. Edler, H. Purnhagen, et C. Ferekidis, "ASAC– Analysis/synthesis codec for very low bit rates", Preprint 4179 (F-6) 100th AES Convention, Copenhagen, 11-14 Mai 1996.

[5] S. Levine, Audio Representations for Data Compression and Compressed Domain Processing, PhD thesis, Stanford University, Août 1998.

[6] H. Purnhagen, N. Meine, "HILN - The MPEG-4 Parametric Audio Coding Tools", IEEE International Symposium on Circuits and Systems (ISCAS 2000), Genève, Suisse, Mai 2000.

[7] E. Schuijers, W. Oomen, B. den Brinker and J. Breebart, "Advances in Parametric Coding for High-Quality Audio", 114th AES Convention, Amsterdam, Mars 2003.

[8] F. Itakura, "Line spectral representation of linear predictive coefficients of speech signals", J. Acoust. Soc. Amer., vol. 57, Supplément no. 1, S35, 1975.

[9] B.den Brinker et F. Riera-Palou, "Pure Linear Prediction", 115th AES Convention, New York, Octobre 2003.

[10] http://www.chiariglione.org/mpeg/working_documents/mpeg-04/audio/param-audio-VT.zip

[11] ISO/IEC JTC1/SC29/WG11/N6428,"ISO/IEC13818-7:2004 (AAC 3rd edition)", Mars 2004, Munich,

[12] 3GPP TS 26.190, "Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Transcoding functions", 2004.

[13] M. Ghenania, C.Lamblin, "Low-cost Smart Transcoding Algorithm between ITU-T G.729 (8 kbit/s) and 3GPP NarrowBand AMR", Eusipco 2004.

[14] ITU Radiocommunication Study Group 6, "Recommandation ITU-R BS.1387 – Method for objective measurements of perceived audio quality"